Historical Biology
An International Journal of Paleobiology

Taylor & Francis
Taylor & Francis Group

Historical Biology
An International Journal of Paleobiology

ISSN: 0891-2963 (Print) 1029-2381 (Online) Journal homepage: www.tandfonline.com/journals/ghbi20

Species discovery and dental ecometrics: good news, bad news and recommendations for the future

Oscar E. Wilson, Mikael Fortelius & Juha Saarinen

To cite this article: Oscar E. Wilson, Mikael Fortelius & Juha Saarinen (2023) Species discovery and dental ecometrics: good news, bad news and recommendations for the future, Historical Biology, 35:5, 678-692, DOI: 10.1080/08912963.2022.2060102

To link to this article: https://doi.org/10.1080/08912963.2022.2060102

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

View supplementary material

Published online: 18 Apr 2022.

Submit your article to this journal

Article views: 1869

View related articles

View Crossmark data

Citing articles: 4 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ghbi20

HISTORICAL BIOLOGY
2023, VOL. 35, NO. 5, 678–692
https://doi.org/10.1080/08912963.2022.2060102

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

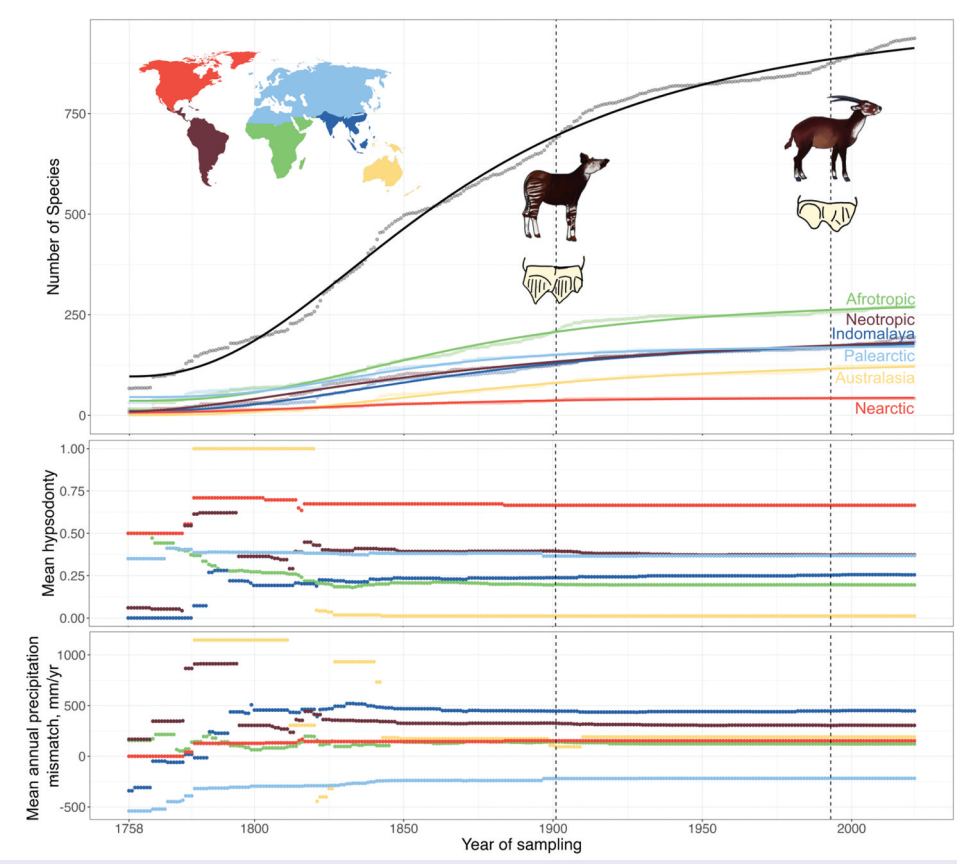
Species discovery and dental ecometrics: good news, bad news and recommendations for the future

Oscar E. Wilson, Mikael Fortelius and Juha Saarinen

Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland

ABSTRACT

New mammal species have always been a source of major zoological intrigue, and whilst discoveries of large mammals are becoming less frequent, species like the saola and dingiso continue to enter the zoological literature. Modern communities are often assumed to be complete and are used for constructing community-level models. One example is ecometric modelling. Here, we estimate the number of large herbivorous mammals not yet described based on description curves. We also investigate the effect of species description on the stability and accuracy of previously established relationships between traits and climate. The contemporary description record of large herbivorous mammals is incomplete, with at least 83 species undescribed. Primates, artiodactyls and marsupials are estimated to contain the greatest undescribed diversity, with particular gaps in the Neotropic, Afrotropic and Indomalayan realms. We find that beyond ~40% completeness, there is a limited impact on trait–environment relationships of increased species description, but that there is a high mismatch between true and predicted climatic values for published models. Consequently, mammalian species that have not yet been discovered are unlikely to have a large impact on the accuracy of trait-environment models, but we suggest possible alterations to previous approaches that might improve the accuracy of future models.



ARTICLE HISTORY
Received 25 February 2022
Accepted 27 March 2022

KEYWORDS
Species discovery; ecometrics; species accumulation curves; cryptozoology; community modelling

CONTACT Oscar E. Wilson oscar.wilson@helsinki.fi Department of Geosciences and Geography, University of Helsinki, Helsinki 00014, Finland
Supplemental data for this article can be accessed here.

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Published online 18 Apr 2022

Introduction

In 1900, Sir Harry Johnston wrote a letter to the Secretary of London's Zoological Society, Dr. Philip Sclater, to inform him of investigations he had been carrying out in the forests of the Democratic Republic of the Congo on an animal referred to as an o'api by the local Wambutti people (Shuker 2012). Johnston's expeditions and negotiations at Fort Mbeni, as well as those of others yielded enough material for a formal species description by Sclater in 1901 as *Equus(?) johnstoni*(Sclater 1901a, 1901b) and for a reclassification as *Okapia johnstoni* (rather than the genus *Helladotherium* preferred by Johnston himself) by E.R. Lankester later that same year (Lankester 1901a, 1901b). This discovery was described as the 'most important zoological find in recent times' (Mill Valley Independent 1909), demonstrating that large mammals might remain undiscovered.

Since its discovery many large mammal species have been described by science, many of which had been previously known to indigenous populations (Shuker 2012). These include the presumed-extinct Chacoan peccary in 1975 (*Catagonus wagneri*) (Wetzel et al. 1975), the saola (*Pseudoryx nghetinhensis)* in 1992 (Dung et al. 1993), dingiso (*Dendrolagus mbaiso*) in 1994 (Flannery and Szalay 1995), kipunji (*Rungwecebus kipunji*) in 2005 (Jones et al. 2005) and as recently as 2021, the Benin tree hyrax (*Dendrohyrax interfluvialis*) (Oates et al. 2021). The frequency of recent discoveries suggests that larger mammals are likely to be discovered in the future.

In many cases, these may represent cryptic species, previously hidden in plain sight but unrecognised as species. Such species are well recognised across animal groups (Pfenninger and Schwenk 2007) and have been previously found in large mammals. One such example is the African forest elephant, *Loxodonta cyclotis*, that was initially described by P. Matschie in 1900 (Matschie 1900). This species was considered by many to be a subspecies of the bush elephant, *Loxodonta africana* until a reassessment in 2000 based on ecology, morphology (Grubb et al. 2000), and subsequent genetic analyses (Groves and Grubb 2000; Roca et al. 2001, 2015). Subsequent analyses using ancient DNA have even suggested a closer affinity to extinct European straight-tusked elephants than *L. africana* (Meyer et al. 2017). Such cases of the revision of subspecies to species status represent a significant method of new species descriptions.

Cryptozoology, defined by Bernard Heuvelmans as 'The scientific study of hidden animals, i.e., of still unknown animal forms about which only testimonial and circumstantial evidence is available, or material evidence is considered insufficient by some' (Heuvelmans 1982), is often considered a pseudoscience because of the biological improbability (or impossibility) of subjects investigated, inconsistent use of nomenclature and often fruitless searches for evidence (Rossi 2016). This pseudoscientific label has made researchers reluctant to publish on potential new species without physical evidence (Burney and Ramilisonina 1998; Oren 2001). Even where physical evidence is provided, the perceived unlikelihood of discovery of new species of large mammals has led to recent academic disputes (e.g., the proposed new species of tapir, *Tapirus kabomani* in 2013 (Cozzuol et al. 2013, 2014; Voss et al. 2014)). Despite these challenges, new large mammals continue to be described, and are likely to continue to represent significant new finds in the field of conservation, ecology and palaeobiology. As we continue to enter the sixth major mass extinction event in Earth's history (Barnosky et al. 2011), dark extinction is likely to become more prevalent, as species become extinct before being described (Boehm and Cronk 2021). Undiscovered species are likely to experience a greater risk of extinction (Giam et al. 2011; Liu et al. 2022) so description may allow for conservation prioritisation, and searches for 'missing' taxa can provide other benefits for conservation, including increased stakeholder engagement and funding (Watson and Davis 2017; Holmes et al. 2018; Padovani and Rossi 2021).

Ecometrics describes a trait-based approach used to understand variability in the environment through the relationship between environmental variables and functional traits within a community . It has been described by some authors as a 'taxon-free' approach (Andrews & Hixson 2014), given the focus on traits themselves, rather than the organisms that hold them (Polly and Head 2015; Vermillion et al. 2018). Functional traits are those that have some impact on organism fitness, through survival, development and growth (Violle et al. 2007). Ecometric approaches have used links between functional traits and their environment to reconstruct the environment in the past and future. Ecometric approaches have applicability in predicting environments both in the past and in the future, as a result of their focus on traits themselves rather than the species that possess them (Barnosky et al. 2017; Vermillion et al. 2018). A variety of functional traits have been identified, across taxonomic groups, including body size in ectotherms as a predictor of mean annual temperature (Head et al. 2009), stomata counts in plants as a predictor of carbon dioxide concentration (Beerling et al. 2002) and the calcaneum gear ratio in carnivorans (Polly 2010). The teeth of large mammalian herbivores have been applied widely to ecometric approaches, due to their strong links to environmental factors (Jernvall et al. 1996; Fortelius et al. 2002; J.T. Eronen et al. 2010a; Damuth and Janis 2011; Liu et al. 2012; Žliobaite et al. 2016; Oksanen et al. 2019) and high preservation potential within the fossil record (Zliobaite and Fortelius 2021).

As cheek teeth in large herbivores are critical for food processing, they represent the interface between the organism and its environment. Previous authors (e.g., Liu et al. 2012, Eronen et al. 2010a; Žliobaite et al. 2018) have described that traits such as hypsodonty and the number of longitudinal lophs on molar teeth have a significant relationship with climatic variables including temperature and precipitation.

Models used by previous studies have been built on contemporary herbivore communities using taxa from Artiodactyla, Perissodactyla, Primates and Proboscidea, and have assumed these modern communities are completely represented. Given the utility of such models in predicting climate and vegetation patterns in the past, as well as in conservation biology (Barnosky et al. 2017 ; Vermillion et al. 2018), these assumptions should be tested. In this study, we assess to what extent this incompleteness affects community-level descriptors, using dental ecometrics as a case study. Organisms are most likely to be described first if they are more common (Patterson 2000; Reeder et al. 2007) and common taxa have been found to provide the greatest link between dental traits and the environment (Jernvall and Fortelius 2002). As a result, we consider that real species description records, rather than randomly sampling communities offer the best opportunity to investigate the stability of ecometric predictions to completeness.

First, using historical description records, we predict how many large mammal species alive today have not yet been described within significant taxonomic groups and within each of the six biogeographic realms to predict where future sampling efforts should be targeted. We then use dental ecometric models from the literature to examine how the stability and accuracy of these models has changed with the increased discovery of large mammals in recent history. Using these results, we discuss whether it remains appropriate to ground these models on recent communities and suggest possible improvements that might be considered.

Materials and methods

Mammal taxonomy

We used the list of mammals available in the Mammal Diversity Database, mammaldiversity.org (accessed 22 October 2021). We then refined the selection to include only large, terrestrial herbivores, defined in this case as mammals that receive the majority (i.e., greater than 50%) of their energy from plant material (estimated from previously published literature), which weigh over 1 kg and spend most of their lives on land. Body mass estimates were taken from the published values in the Pantheria database (Jones et al. 2009) and the mean of the minimum and maximum mass given by Freudenthal and Martín-Suárez 2013. Species were included if estimates from both databases exceeded 1 kg. Where data were unavailable for a species, we estimated whether it weighed over 1 kg from the records for congenerics. This left a total of 930 mammal species from 12 orders (Artiodactyla, Carnivora, Chiroptera, Dermoptera, Diprotodontia, Hyracoidea, Lagomorpha, Perissodactyla, Pilosa, Primates, Proboscidea, Rodentia) (Supplementary Material 1). We used the *authoritySpeciesYear* recorded in the Mammal Diversity Database as the date of first description and *biogeographicRealm* for location information. We acknowledge that this is the date of description, which may be considerably later than the date of discovery, though this has the advantage of removing an element of ambiguity, as description dates tend to be the 'best documented' of the dates between the discovery of the first specimens of a species and formal recognition by the scientific community (Pine 1994).

Species discovery models

We fitted a logistic regression curve to species description records as had been done previously for pinniped description records (Woodley et al. 2009). The input variables were a numeric value for the year of description (ranging from 0 to 263 to correspond to records from the years 1758 to 2021) and the cumulative number of species described by that year. A logistic regression curve showed the closest match to the description records on visual inspection of all models tested (Supplementary Material 3). We fitted curves using the '*drc*' package (Ritz et al. 2015) on RStudio 2021.09.0.351 (RStudioTeam 2021). We fitted curves separately for a description record of all large mammals and for six groups of herbivores (Artiodactyla, Marsupialia, Paenungulata, Perissodactyla, Primates, Rodentia). We also separately fitted curves for each biogeographic region included in the dataset (Afrotropic, Australasia, Indomalaya, Nearctic, Neotropic and Palearctic). We used a four-parametric log logistic regression model with the function *LL.4()* in *drc*.

$$f(x) = c + \frac{d - c}{1 + exp(b(log(x) - log(g)))}$$

This model has 4 parameters, where b gives the slope, c the lower limit, d the upper limit and g the midpoint about which the curve is symmetrical. The *drc* package estimates the residual standard error for each model and provides p values for each parameter using the *summary()* function. When fitting the model for Australasia only, the estimate for c was found not to be significant at the level of $p = 0.05$. In this case, we fixed the value of c at 0 with the *LL.3()* function. To find the estimated number of undiscovered species, we used the value of d, the upper limit.

We reran the models for each biogeographic realm including only the taxa that have previously been used in dental ecometrics studies (Artiodactyla, Perissodactyla, Primates, Proboscidea) (henceforth called 'Ecometric Taxa') to give an estimate of how many of these are currently undiscovered in each realm.

Late pleistocene diversity

We downloaded species-level records of large herbivores from the Palaeobiology Database (The Paleobiology Database 2021). We searched for all species from the start of the Late Pleistocene (0.129Mya)(Negri et al. 2015) to the recent. Because these records were then compared with the ecometric models constructed using previous methods (Liu et al. 2012; Oksanen et al. 2019), our search included only species from Artiodactyla, Perissodactyla, Primates and Proboscidea. We obtained estimates for the number of species missing from a given biogeographic realm by comparing the list to the present and removing species that are still found in the same biogeographic realm.

Dental ecometric traits

Previous studies using dental ecometric traits have included only taxa from Artiodactyla, Perissodactyla, Primates and Proboscidea (Liu et al. 2012; Žliobaite et al. 2016, 2018; Oksanen et al. 2019; Saarinen et al. 2021). We therefore removed all other taxa from the original list of 930 large, herbivorous mammals, giving a list of 604 ecometric species (65%). For each species, we recorded three traits, previously used to estimate climate – hypsodonty (HYP), number of longitudinal lophs (LOP) and number of acute lophs (AL) (Full description of coding scores and functional relationships available in Supplementary Material 2). Values for these dental traits were taken from the codings by Oksanen et al. 2019. Where no codings were available for a particular species, the values for the three dental ecometric traits were estimated from values for other species within the genus, as intrageneric variability in these traits is low. Where a genus included species with different values, we estimated the value from the closest phylogenetic relative. For monospecific genera (*Pseudoryx* and *Rungwecebus*), values for the dental traits were estimated from publicly available photographs.

Ecometric models

For all ecometric analyses, we removed domesticated taxa, as prior analyses suggested there was negligible effect on the trends shown (see Supplementary Material 6). We subsetted the list of mammalian herbivores from the four taxa previously used in ecometric studies by year of discovery, to create a list of mammals with the three dental traits recorded for each year between 1758 and 2021. We used this approach, based on real species description curves, because the timeline of description is likely to correlate with commonness (Reeder et al. 2007). Common animals are more likely to capture the trait-environment signal (Jernvall and Fortelius 2002) and therefore are critical for ecometric modelling.

Range data for the mammals were taken from the IUCN Red List, available at https://www.iucnredlist.org/resources/spatial-data-download. We divided the Earth into a series of hexagonal grid cells 50 km across and took the centroid of each of those grid cells. The centroids were intersected with the biogeographic realm polygons from Ecoregions of the World 2017 (M. Mechenich, unpublished raw data, 2021). For each centroid, the mean annual temperature (MAT) and mean annual precipitation (MAP) were

extracted using the '*sp*' (Pebesma & Bivand 2005; Bivand et al. 2013) and '*raster*' (Etten et al. 2021) packages in RStudio from the WorldClim database (Fick and Hijmans 2017).

Following Vermillion et al. 2018, we found the mean value for each dental trait for the community at each sampling point for each year between 1758 and 2021. We included only sampling points where richness was greater than 5 species (Vermillion et al. 2018). We then calculated the estimated MAT and MAP using existing dental ecometric models for large herbivores. Liu et al. 2012 used HYP and LOP.

$$MAT = 24.7 + 13.8HYP - 25.1LOP$$

$$MAP = 2727.7 - 411.9HYP - 859.7LOP$$

Oksanen et al. 2019 instead used AL to estimate MAT and a combination of HYP and LOP to estimate MAP (Equations 4 and 5).

$$MAT = 27 - 28.5AL$$

$$MAP = 2491 - 289HYP - 841LOP$$

Oksanen et al. 2019 suggested that the models in Liu et al. 2012 should perform better (i.e., have more similarity to the true MAT and MAP) in the tropics, whilst the equations from Oksanen et al. should be more accurate in temperate regions. For both existing models and for every year between 1758 and 2021, we calculated the mismatch between the true and modelled values for both MAT and MAP, by subtracting the modelled value from the 'true' values extracted from the WorldClim database at every centroid. These 'true' values represent the climatic condition in 2021, rather than for each year studied. We plotted the mean trait scores for each dental trait, and the mismatches in both MAT and MAP for each centroid to identify any patterns between the two. To test for differences between the biogeographic realms, we averaged the trait scores and mismatch in both MAP and MAT across all points in each realm for each year. We produced all plots using the *ggplot2* package (Wickham 2016).

To assess the impact of sampling order on the results, we also repeated the ecometric analyses after randomly resampling the description dates for all taxa (Supplementary Material 8).

Results

Species description rates per clade

The estimated model parameters for each group and for all large mammals are listed in Table 1. The upper limit (*d*) in the log-logistic model gives the maximum number of species expected to be discovered according to prior discovery rates. The log-logistic model for all large herbivorous mammals gave an estimate of 83 species not yet discovered. This is lower than the sum of the other sub-groups of large mammals for which we obtained estimates, suggesting some uncertainty in this result.

The clade with the largest number of species not currently described is Primates (64), then Artiodactyla (22) and Marsupialia (17). The expectation from this model is that there are no extant paenungulates not currently described, and possibly only a single perissodactyl species. This result matches the expectation from visual inspection of the model for each group (Figure 1) where Perissodactyla and Paenungulata are the only two groups that appear to have reached an asymptote for their rates of description.

Species description rates by biogeographic realm

As with the models grouped by taxonomy, we found that the number of species predicted to be discovered using all realms (83) was lower than the sum of the results obtained from each realm modelled separately (103). The realms with the lowest number of species left to be described are expected to be the Palearctic (1) and the Nearctic (4). In contrast, the Neotropic has the highest number of species left to be described according to our model (32), followed by Afrotropic (23) (Figure 2, Table 2).

Number of ecometric taxa left to be described by realm

Including only the taxa previously used in dental ecometric studies showed that Indomalaya was the biogeographic realm expected to have the highest number of species left to be described (29), followed by Neotropic (23) and Afrotropic (21). Australasia, Nearctic and Palearctic each only have a single species left to be described from within the clades previously used for dental ecometric studies (Table 3).

Missing late pleistocene fauna

The different biogeographic realms show significant differences in the completeness of their large herbivore faunas after the End Pleistocene megafaunal extinctions. The Nearctic is missing the largest percentage of taxa from groups that have historically been used for ecometric analyses (78.79%). Australasia is not missing any of these species, whilst the Afrotropic and Indomalayan realms are missing comparatively low number of species, at (10.66% and 9.43% respectively). Whilst the fauna of the Neotropic realm was dramatically impacted by the End-Pleistocene extinction (Barnosky et al. 2004), it is only missing 25% of the taxa that have historically been used for dental ecometric analyses because most of the lost South American herbivores were from other clades whose dental ecometric relationships have not yet been studied (e.g., Litopterna, Notoungulata, Chlamyphoridae, Pilosa).

Table 1. Estimated model parameters for the log-logistic model used to predict the number of living large herbivorous mammals not yet described for each taxonomic group based on prior rates of description. The slope of the model is given by b, the lower limit by c, the upper limit by d and the midpoint by g. For each model, the standard error (SE) is given. The number of organisms described by 2021 is listed, and the future species to be described is the difference between the value of d and the number in 2021. Model set 1758 as equal to 0. All models used were a four-parameter log-logistic regression model.

Taxon	b	c	d (± SE)	g	SE	2021 number	Future species
Total	2.33	95.76	1012.50 (± 9.73)	109.77	18.77	930	83
Artiodactyla	2.33	47.49	281.10 (± 2.92)	95.87	6.95	259	22
Perissodactyla	3.31	6.63	18.40 (± 0.08)	66.8	0.56	18	0–1
Primates	1.97	16.12	385.84 (± 6.48)	129.29	7.09	329	57
Rodentia	2.63	13.31	134.29 (± 1.45)	111.24	3.35	127	7
Paenungulata	3.87	2.04	8.45(± 0.08)	101.28	0.41	9	0
Marsupialia	2.3	−1.37	122.63 (± 2.17)	126.68	3.25	106	17

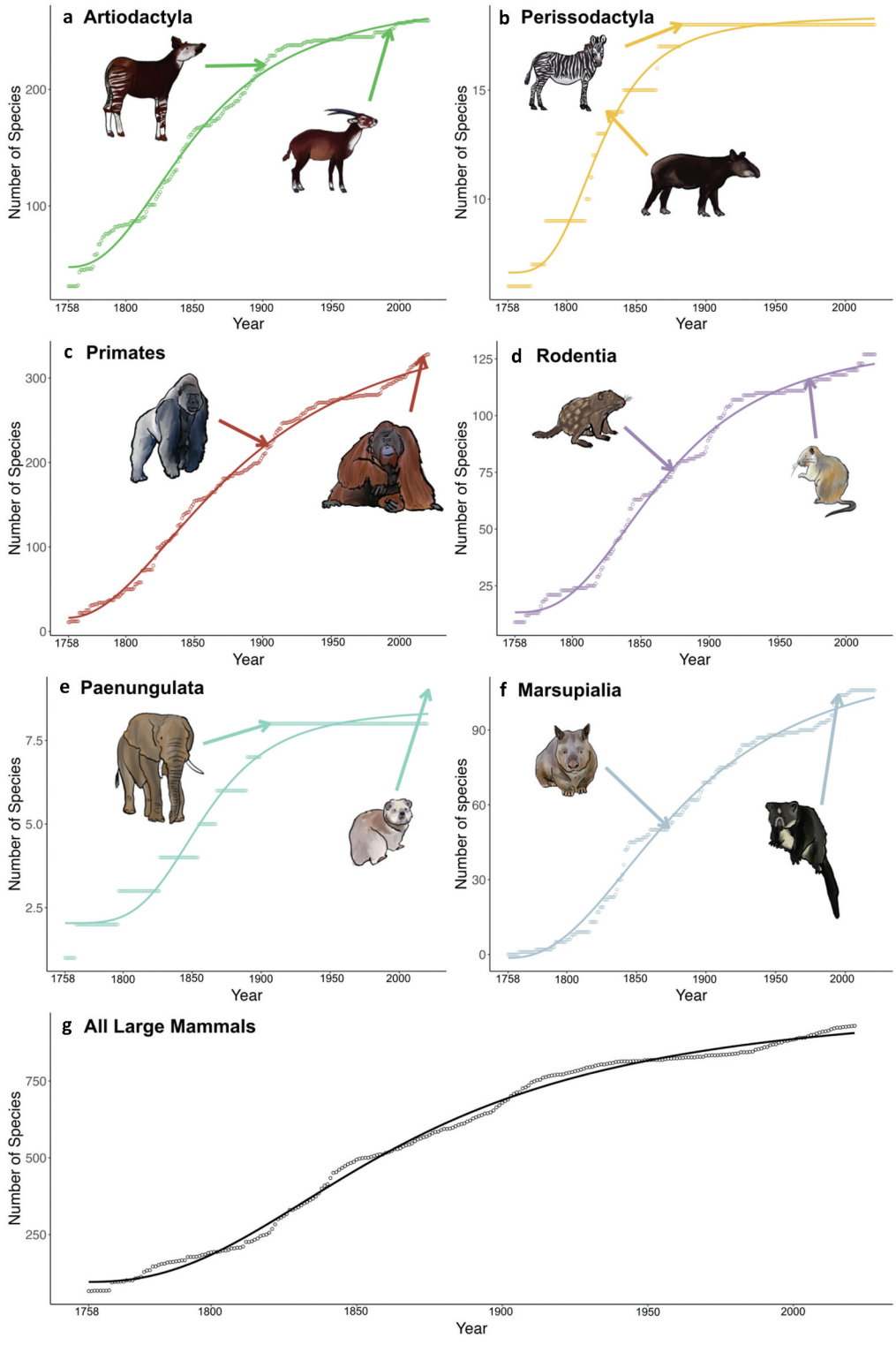


Figure 1. Log-logistic curves fitted to the species description records for large herbivorous mammal groups. (A) Artiodactyla (B) Perissodactyla (C) Primates (D) Rodentia (E) Paenungulata (F) Marsupialia (G) All large mammals. Each point represents the cumulative number of species described by that year, and the filled line is the result from the log-logistic model. Representative species for each group with a significant history of zoological description are labelled. From top-left to bottom-right: Okapi (*Okapia johnstoni*), saola (*Pseudoryx ngethinhensis*), Grevy's zebra (*Equus grevyi*), mountain tapir (*Tapirus pinchaque*), mountain gorilla (*Gorilla beringei beringei*), Tapanuli orangutan (*Pongo tapanuliensis*), pacarana (*Dinomys branickii*), Garrido's hutia (*Capromys garridoi*), Benin tree hyrax (*Dendrolagus interfluvialis*), African forest elephant (*Loxodonta cyclotis*), dingiso (*Dendrolagus mbaiso*), Northern hairy nosed wombat (*Lasiorhinus krefftii*). Illustrations by Vanessa Williams.

Ecometric models

All three dental ecometric traits showed similar patterns, with instability in the average values across each realm until the early 19th century. Mammals that were described after this time had limited impact on the mean trait values (Figure 3). Africa and Eurasia showed the greatest variability in each dental trait (Supplementary Material 4). Mean hypsodonty was highest in northern Africa, southern Eurasia, and the southern tip of South

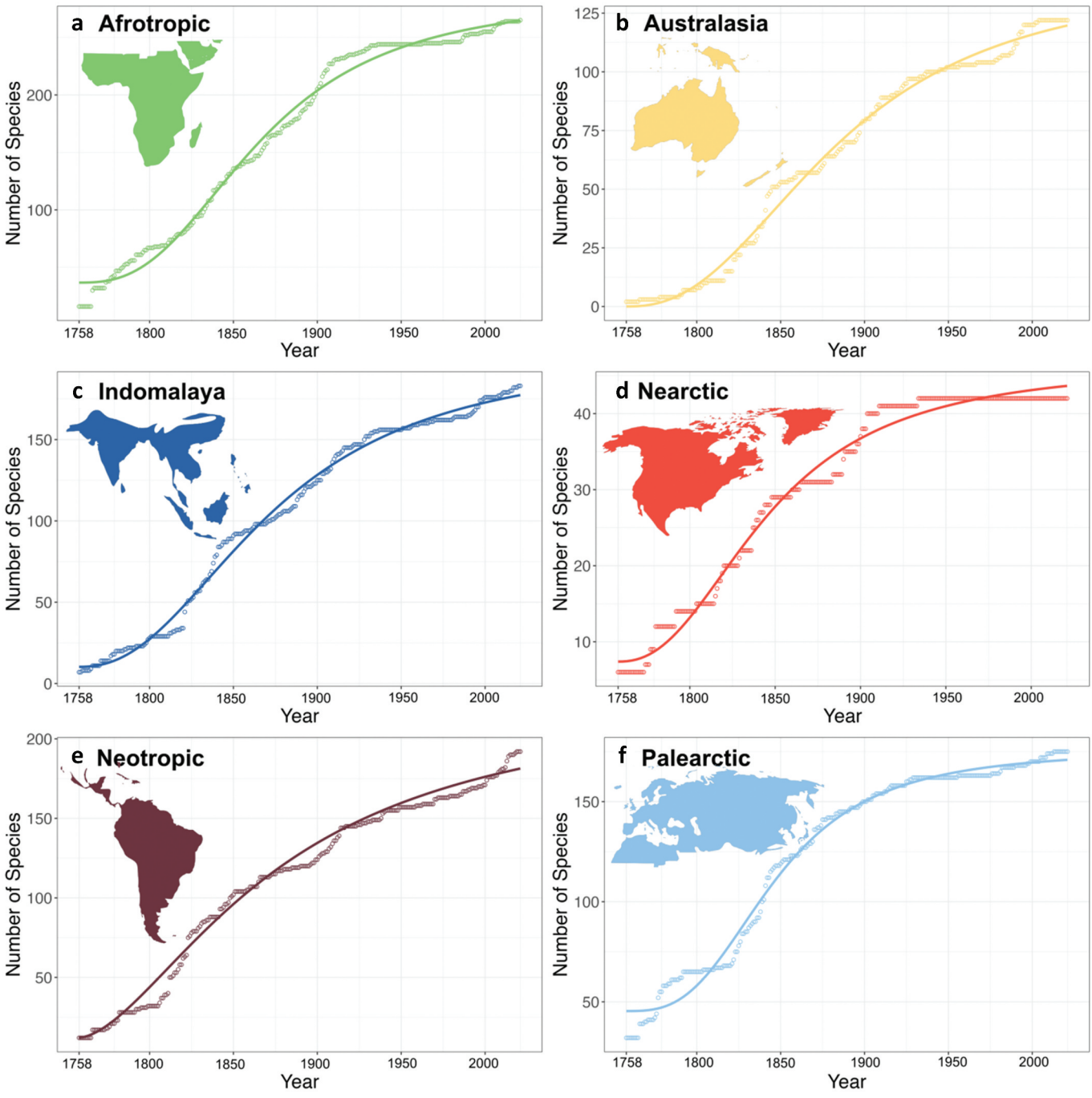


Figure 2. Log-logistic curves fitted to the species description records for each biogeographic realm (A) Afrotropic (B) Australasia (C) Indomalaya (D) Nearctic (E) Neotropic (F) Palearctic. Each point represents the cumulative number of species described by that year, and the filled line is the result from the log-logistic model. Silhouettes by Hal Wilson.

America (Figure S8). Communities in high-latitude environments and those in northern Africa had the highest mean number of longitudinal lophs (Figure S9). Communities in North America and northern Eurasia had a high proportion of species with acute lophs (Figure S10). These geographic trends for all traits were persistent between 1758 and 2021 (Figures S8–10).

Using both models and for all regions, the estimates obtained for both MAT and MAP match the patterns we found in the trait values. Year is a proxy for completeness, as the completeness of each realm's communities increases over time. The estimates are initially inconsistent and changed as more species were described (Figure 4) (Supplementary Material 7). From ~1850 onwards (corresponding to ~40%), the estimates from all biogeographic realms no longer change with increased sampling. The concordance between the patterns observed using trait values and both ecometric models suggests that the stability beyond this time is due to the completeness of the communities, rather than any intrinsic property of either model. The large differences between the mean modelled values and the true values suggest that other factors beyond completeness affect the efficacy of prediction.

Table 2. Estimated model parameters for the log-logistic model used to predict the number of living large herbivorous mammals not yet described for each biogeographic realm based on prior rates of description. The slope of the model is given by b, the lower limit by c, the upper limit by d and the midpoint by g. For each model, the standard error (SE) is given. The number of organisms described by 2021 is listed, and the future species to be described is the difference between the value of d and the number in 2021. Model set 1758 as equal to 0.

Region	Model parameters	b	c	d (± SE)	g	SE	2021 number	Future species
Total	4	2.33	95.76	1012.50 (± 9.73)	109.77	18.77	930	83
Afrotropic	4	2.67	36.77	286.37 (± 3.42)	109.04	8.22	263	23
Australasia	3	2.39	0.00	141.16 (± 2.25)	127.97	3.75	122	19
Indomalaya	4	2.28	10.26	203.46 (± 0.36)	116.55	4.77	185	19
Nearctic	4	2.34	7.39	46.50 (± 0.52)	88.74	1.53	42	4
Neotropic	4	1.67	9.36	224.19 (± 5.68)	119.99	5.45	192	32
Palearctic	4	2.98	45.38	175.76 (± 1.35)	88.61	5.14	175	1

Table 3. Estimated model parameters for the log-logistic model used to predict the number of living large herbivorous mammals included within taxa previously used for dental ecometric studies (Artiodactyla, Perissodactyla, Primates, Proboscidea) not currently described for each biogeographic realm based on prior rates of description. The 2021 ecometric refers to the number of species from the four taxa previously used in ecometric studies and future ecometric is the difference between this value and the estimated total (d). The 2021 non-ecometrics are the number of species that do not belong to any of the four previously used taxa that have been described in each realm by 2021. The slope of the model is given by b, the lower limit by c, the upper limit by d and the midpoint by g. For each model, the standard error (SE) is given. The number of future ecometric taxa is given by the difference between the number in 2021 and the estimate of the upper limit of the model, d.

Region	b	c	d (± SE)	g	SE	2021 ecometric	Future ecometric	2021 non-ecometric	% non-ecometric taxa (2021)
Afrotropic	2.53	32.75	252.67 (± 3.41)	108.46	7.47	232	21	31	11.78
Australasia	3.18	0.78	12.30 (± 0.21)	18.66	0.67	11	1	111	90.98
Indomalaya	2.06	4.90	171.97 (± 3.50)	135.07	3.77	143	29	42	22.70
Nearctic	1.78	2.59	14.99 (± 0.14)	53.20	0.53	14	1	28	66.67
Neotropic	1.55	4.33	148.76 (± 5.21)	118.34	4.656	126	23	66	34.44
Palearctic	2.66	32.10	113.34 (± 1.12)	82.36	3.80	112	1	63	36.00

Both previously published dental ecometric models showed differences between the true MAT and MAP and the predictions, though these mismatches were different for each biogeographic realm (Figure 4, Table 4). Both models predicted southern Eurasia to be hotter and wetter than the true value (Supplementary Material 5), which is probably associated with the high value of hypsodonty in this region (Figure S8). The east coast of North America was estimated to be cooler and drier than the true value (Figures S11–S14). These mismatches are geographically correlated with the communities that are brachydont, with a high proportion of acute lophs and high number of longitudinal lophs. The temperature estimates based on Oksanen et al. 2019 (Figure S11) gave a much warmer value for high latitudes in the Northern Hemisphere than the true values and compared to Liu et al. 2012 (Figure S13). Communities in these areas contain almost no acute lophs (Figure S10).

No estimates could be made for the MAT and MAP for Australia or New Guinea because there are no taxa that have previously been used for ecometric studies within these islands.

Discussion

How many large mammals are left to describe?

Our log-logistic models suggest that there are likely to be at least 83 species of large (>1 kg) mammals left to describe worldwide, though the true number may be higher, given the sum of the predictions from only the few select taxonomic groups examined here is higher (103). We acknowledge that previous studies have found that estimation of predicted numbers using discovery curves can be associated with a large margin for error (Bebber et al. 2007), but given the relative completeness of our understanding of large mammals, we consider our estimates to be reliable.

The number of undescribed large mammals strongly differs between taxa. We estimate that there are a relatively large number of primates, artiodactyls and marsupials not yet described, with very few (if any) perissodactyls and paenungulates (Table 1). However, the reappraisal of the African forest elephant (Groves and Grubb 2000; Grubb et al. 2000; Roca et al. 2001, 2015), discovery of *Dendrohyrax interfluvialis* in 2021 (Oates et al. 2021) and the recognition of unknown vocalisations by hyraxes in the Taita Hills that may be suggestive of a new species (Rosti et al. 2020) suggest that the situation for hyraxes may be more complex than previously recognised. Similarly, whilst *Tapirus kabomani* (Cozzuol et al. 2013) (possibly a synonym of *Tapirus pygmaeus* (van Roosmalen 2013a, 2014) is currently viewed as a subspecies of *T. terrestris* and further investigation is required to determine the nature of *T. terrestris* in relation to both *T. kabomani* and *T. pinchaque* (Voss et al. 2014).

Many of the recent descriptions of large mammals involve redesignation of populations into distinct species. Orangutans provide an example of this, with an original separation between the Borean (*Pongo pygmaeus*) and Sumatran (*P.abelli*) orangutans based on genetic evidence in 1996 (Xu and Arnason 1996) and then in 2017 the designation of *P. tapanuliensis* as a distinct species (Nater et al. 2017). Species description based solely on genetic evidence is a flawed approach, which relies on the availability of technology, and may not be comparable with historical taxonomic approaches (Zamani et al. 2022). Phylogenetic species concepts using both monophyly and diagnosability are generally positive, but rely on availability of specimens and in the past, taxonomic assessments of mammals using phylogenetic concepts have not been transparent (Gutiérrez and Garbino 2018). A consistent, widely applied approach to species delimination would reduce conflicting views on potential species-subspecies delimination. A 'traffic-light' taxonomic method has been employed for felid taxonomy, whereby species certainty is determined based on morphological, genetic, biogeographic, behavioural, ecological and reproductive criteria (Kitchener et al. 2007, 2022). Use of this approach altered the understanding of species number in fields from historical estimates (Wozencraft 2005) and application to other large mammal groups will likely prove useful in cases where recognition of species status is controversial, for example, giraffes (Groves and Grubb 2011; Fennessy et al. 2016), tapirs (Cozzuol et al. 2013, 2014; Voss et al. 2014) and rhinos (Groves et al. 2010; Groves and Grubb 2011).

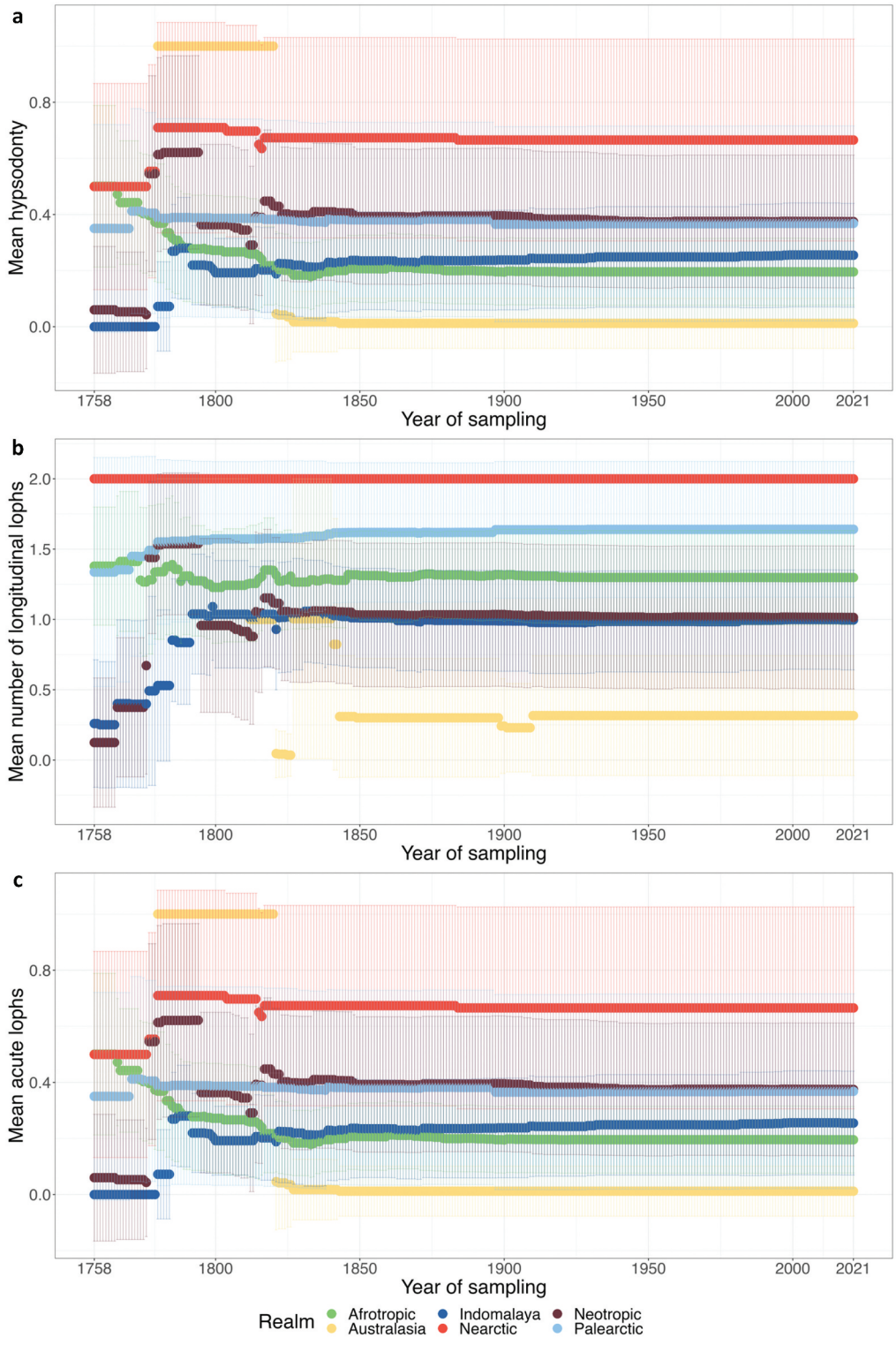


Figure 3. Mean trait values for (A) hypsodonty, (B) longitudinal loph count and (C) acute lophs for every year between 1758 and 2021 where all domesticated taxa are included. Year is a proxy for completeness, given the continued description of mammal species over time. Mean hypsodonty score refers to the average ordination from 1–3. Mean longitudinal loph count is the mean number of longitudinal lophs in the realm and the mean acute lophs is the proportion of large herbivores in the realm with acute lophs. The shaded areas represent ± one standard deviation from the mean.

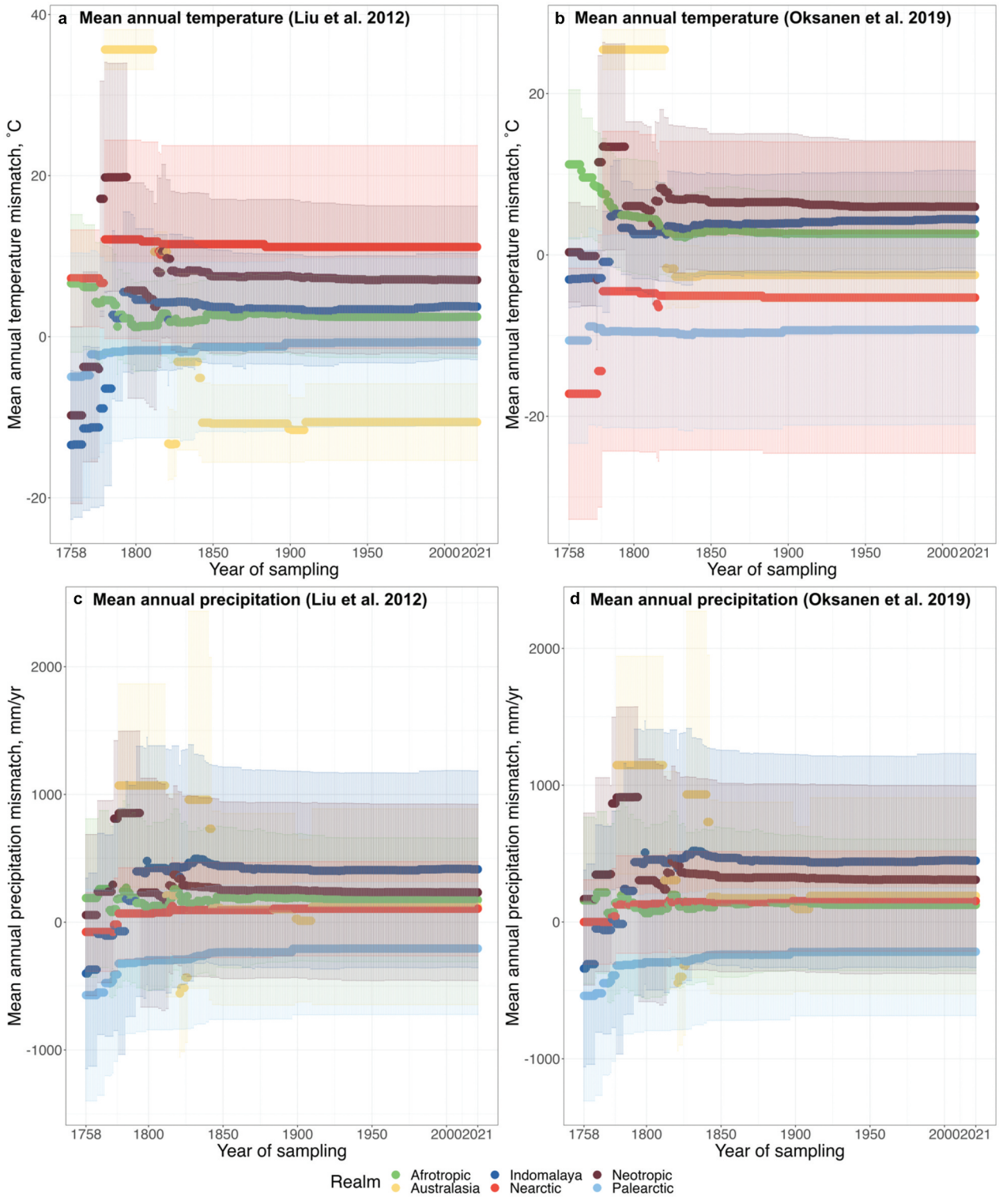


Figure 4. Mismatch between the true mean annual temperature (MAT)(A,B) and mean annual precipitation (MAP)(C,D) in each year using the previously published ecometric models by Liu et al. 2012 (A,C) and Oksanen et al. 2019 (B,D). Each point represents the mean mismatch for that biogeographic realm in that year across centroids. Year is a proxy for completeness, given the continued description of mammal species over time. The mismatch was calculated by subtracting the modelled value from the true value at each centroid. This means that a negative value is suggestive that the modelled value was too hot/wet and a positive value too cold/dry. The shaded areas represent ± one standard deviation from the mean.Four plots showing the mismatch between the actual and modelled mean annual temperature and mean annual precipitation by biogeographic realm for two previously published models (Liu et al. 2012; Oksanen et al. 2019). In each plot, there is initial instability in the average mismatch between the two with increasing completeness of the description records, until around 1850, at which point the mismatches are stable to the increasing number of taxa. In all cases, however, there are multiple realms with a large difference between the true and modelled values

In addition, we consider it likely that large mammals will continue to be for the first time based on new discoveries, as for the new hyrax species (Rosti et al. 2020; Oates et al. 2021). Cryptozoological and zooethnological evidence are likely to be critical here, as discoveries of previous large herbivores by science have been supported by local knowledge (e.g., the dingiso (*Dendrolagus mbaiso*) is considered by the Moni people of Irian Jaya to be their direct ancestor and was extremely well understood by several local tribes prior to its description in 1995 (Flannery and Szalay 1995)). There are many possible discoveries based on similar accounts, with at least 48.1% of primate species discovered since 2000 known by locals prior to their scientific description (Rossi et al. 2018). The 'mapinguari' of Amazonia, which is said to stand >2 m tall, possess coarse brown fur, long claws, an ability to alternate between bipedality and quadrupedality and a terrible smell has been tentatively proposed as a remnant megalonychid ground sloth, for example (Oren 2001; Velden 2016). Other identities, such as Andean bears, are plausible, but ground sloths were present across a variety of habitat types in the Cenozoic, including tropical forests (Pujos et al. 2016) so the faint possibility remains of a remnant population in the Amazon. Similarly, the kilopilopitsofy of Madagascar has been identified as a possible late-surviving hippopotamus given testimony from local Malagasy people matching descriptions of a hippopotamus (Burney and Ramilisonina 1998). Such claims are widespread for large mammals globally (Shuker 2016) and more interdisciplinary collaboration between ethnozoologists, cryptozoologists and biologists, with increased use of traditional ecological knowledge, may help to uncover other possible missing species as a conservation priority (Padovani and Rossi 2021).

Where are we likely to find new species?

Our models suggest that different biogeographic realms are likely to contain different numbers of undescribed species (Table 2). We predict that the Neotropic (32) and the Afrotropic (23) realms have the most species left to describe, with the Palearctic (1) and Nearctic (4) have the fewest. When including only taxa previously used in ecometric studies (Artiodactyla, Perissodactyla, Primates and Proboscidea), we predict that Australasia, Nearctic and Palearctic each have only a single species yet to be described (Table 3). In the case of Australasia, this reflects the fact that only relatively few species from these groups live in this realm, with most niches occupied by marsupials. Otherwise, we can see that the biogeographic realms typified by temperate regions are more completely sampled than those which contain tropical rainforests and exist at lower latitudes.

We suggest two key explanations for this fact. First, lower latitude environments generally contain higher species diversity across taxonomic groups, with species diversity showing a latitudinal biodiversity gradient (LBG), maintained over deep time (Mittelbach et al. 2007; Rolland et al. 2018). In large mammals, the exact explanation for this differs across taxonomic groups, but relates to either a higher rate of speciation, a lower rate of extinction or the combined effect of the two (Rolland et al. 2014). Ultimately, this is due to the increased input of energy from the Sun to the tropics relative to high-latitude regions (Brown and Svenning 2014). The higher total number of species in the tropics means that complete sampling is more challenging, particularly as range sizes are smaller at low latitudes according to Rapoport's Rule (Rapoport 1982; Stevens 1989), reducing the chance of encounter.

In addition, geographic sampling bias has had a strong impact on the description records from tropical and temperate regions. This is true not just for taxonomic research (Menegotto and Rangel 2018), but also in ecosystem services (Clarke et al. 2017) and palaeoecology for example (Vilhena et al. 2013; Stroud and Feeley 2017). As noted by Collen et al., 83% of the records in the Global Biodiversity Information Facility (GBIF) come from only 10 countries, of which only one is tropical (Costa Rica) (Collen et al. 2008). This bias is related to the fact that most researchers live in temperate areas (particularly Europe and North America) and most cooperation between researchers occurs at this high-latitude belt (Leydesdorff and Persson 2010). In palaeontology, the geographical bias in sampling as a result of colonial history, English proficiency and a culture of 'parachute science', among other factors has created a heterogeneous level of biodiversity sampling (Raja et al. 2021). We expect this to be the same for modern biodiversity and continued efforts to sample the tropics through collaboration with local researchers should be prioritised.

Some of the most celebrated 20th- and 21st-century discoveries have been made in specific, isolated tropical regions, including the Laotian rock rat (*Laonastes aenigmamus*) (Jenkins et al. 2005), saola (*Pseudoryx nghetinhensis*) (Dung et al. 1993) and giant muntjac (*Muntiacus vuquangensis*) (Schaller and Vrba 1996) in the Annamites within the Indomalaya realm and the collection of potentially new mammals, some of which are yet to be formerly described and many of which are subject to scrutiny, reported by Marc van Roosmalen from the Rio Aripuanã basin in the Amazon (Shuker 2012; van Roosmalen 2013b). Such areas of apparent endemism in the tropics warrant further investigation and consider it likely that future exploration might produce new species from these and other such tropical rainforests.

Species discovery and ecometric modelling

Using the reported models from both Liu et al. 2012 and Oksanen et al. 2019, we found that there was widespread mismatch between the true values of climatic variables and the modelled values. For example, southern Eurasia was modelled as being too warm and too wet and southern North America as being too cold (Supplementary Material 5). In addition, the models of Oksanen et al. 2019 predicted a temperature that was much too hot in the highest latitude regions of the Northern Hemisphere (Figures S13, 14). Excluding Australasia (which contained only 111 sites), the realm with the greatest temperature mismatch in 2021 using the Liu et al. 2012 model was Nearctic (+11.14 ± 12.59°C) and for the models of Oksanen et al. 2019, the greatest MAT mismatch was for the Palearctic (−9.26 ± 11.75 °C) (Table 5). For both models, the greatest mismatch in MAP was for Indomalaya, where dental ecometrics predicted a much drier climate than is true (Table 5). The geographic patterns for the mismatch between the true and predicted climate correlated with the distribution of the three dental traits used here (Supplementary Material 4, Supplementary Material 5).

Table 4. Number and percentage of species from groups previously used for dental ecometric studies missing from each biogeographic realm since the Late Pleistocene (defined as 0.126Mya). These values therefore include species lost during the End-Pleistocene megafaunal extinction and any extinctions in the Holocene. Species lists for the Late Pleistocene were taken from the Palaeobiology Database. The percentage of missing taxa is equal to the number of missing Late Pleistocene species divided by the sum of the number of species 2021 and the number that are missing.

Realm	2021 ecometric	2021 Ecometric taxa	Missing End-Quaternary Ecometric taxa	% Missing End-Quaternary Ecometric taxa
Afrotropic	232	237	29	10.66
Australasia	11	11	0	0.00
Indomalaya	143	143	15	9.43
Nearctic	14	14	52	78.79
Neotropic	126	127	42	25.00
Palearctic	112	112	48	30.00

Table 5. Mean and SD of mismatches for mean annual temperature (MAT) and mean annual precipitation (MAP) by biogeographic realm for models constructed by both Liu et al. 2012; Oksanen et al. 2019 using taxa described by 2021. N_{2021} refers to the number of centroids containing 5 or more species from taxa previously used in ecometric analyses (Artiodactyla, Perissodactyla, Primates, Proboscidea).

		Liu et al. 2012 mismatch				Oksanen et al. 2019 mismatch			
		Mean annual temperature, °C		Mean annual precipitation, mm/yr		Mean annual temperature, °C		Mean annual precipitation, mm/yr	
Realm	N_{2021}	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Afrotropic	8205	2.47	5.03	174.30	483.98	2.62	5.24	124.40	479.88
Australasia	111	−10.60	4.76	120.67	766.52	−2.49	3.34	191.22	716.83
Indomalaya	3258	3.73	6.58	414.09	770.54	4.41	6.04	447.70	781.30
Nearctic	7645	11.14	12.59	105.45	368.79	−5.29	19.29	152.50	365.64
Neotropic	6930	7.04	9.17	232.96	690.08	5.97	8.14	308.20	686.83
Palearctic	19,612	−0.67	10.37	−206.32	516.60	−9.26	11.75	−217.66	466.98

We found that there is a threshold beyond which an increase in the number of species discovered has minimal effect on the trait averages (Figure 3) or climatic mismatches (Figure 4). This threshold differed for the realms and for the different models, though in general, once >40% of species had been described (Figure S24), there was a limited change with increased community completeness (Supplementary Material 7).

We consider it likely that this pattern is controlled by the fact that more recently described large mammals are likely to have more restricted ranges, be less abundant within an ecosystem, be more likely to be specialist and have a lower body mass (although this is not always necessarily the case)(Patterson 2000; Reeder et al. 2007) (Supplementary Material 8). The dental traits of the most common species within a community tend to be the best predictor of climate (Jernvall and Fortelius 2002), so it matches our expectations that once the most common species have been described, further sampling has only a limited impact on the predictions of climate from dental traits. In addition, the discovery of cryptic species based on the use of genetic evidence means that different species are likely to have almost identical dental functional morphology. Whilst our estimates suggest that there are still many undescribed modern large herbivores, community descriptors are stable to this incompleteness, and for the purpose of ecometric modelling we can continue to assume modern large mammal communities are functionally complete.

The large difference between the true values for the climatic variables and the estimates from our models suggests that the incompleteness of the record used to construct models is not the largest challenge for obtaining accurate climatic estimates from dental traits. The differences between the two models sampled here (Figure 4) suggests that model choice is critical. We also believe that models built on smaller regional or continental scales may produce more accurate estimates, because this will allow for greater optimisation of parameters, and potential inclusion of other taxa within the focal area with a functional effect on vegetation.

The loss of megafaunal taxa at the end of the Pleistocene and through the Holocene is likely to have had a large impact on the accuracy of these models. The Nearctic is the realm with the highest number of lost taxa historically used in ecometric studies due to this extinction event (Table 4) and the greatest mismatch in the MAT estimate (although the estimates for MAP are closest to the true values), which may suggest that specific values of lophedness or number of acute lophs were lost during the extinction event. For example, the estimated temperature for North America is too low using the model of Oksanen et al. 2019 (Figure S13), suggesting contemporary North America contains an overabundance of taxa with a high number of acute lophs (Figure S10). The extinction event at the end of the Pleistocene was generally biased towards open-habitat specialists, with lower rates of extinctions in forest species (Johnson 2002). The modern North American megafauna is depauperate in taxa such as horses and proboscideans that lack such lophs and instead contains a high number of deer, which possess these acute lophs (Oksanen et al. 2019). In addition to global extinction of North American taxa, this continent has seen changes in the climatic ranges of many taxa as a consequence of anthropogenic activity (Pineda-Munoz et al. 2021), which we suggest has consequences for ecometric modelling based on modern ranges of surviving taxa.

According to the Mammal Diversity Database, there were 930 recognised extant species of large herbivorous mammals globally in 2021, 607 of which were from one of the 4 taxonomic groups previously used for ecometric modelling. This leaves a gap of 323 large, herbivorous mammal species from other taxa. In all biogeographic realms, a significant proportion of the large mammalian herbivore species are not from these four taxonomic groups (Table 3). The lowest proportion is in the Afrotropic realm, where only 11.78% of species are non-ecometric taxa. Previously, ecometric models for large mammals have most typically been applied to Plio-Pleistocene environments as a tool to understanding environments in the early evolution of *Homo* (Meloro & Kovarovic 2013; Fortelius et al. 2016; Saarinen et al. 2016, 2021; Žliobaite et al. 2016, 2018; Oksanen et al. 2019) and in other cases have focussed on a specific subset of taxa within a continental or regional scale (Fulwood 2020; Foister and Felice 2021; Short et al. 2021a). Artiodactyls, perissodactyls, primates and proboscideans are the most common large herbivorous mammal taxa in Africa and Southern Europe in the Plio-Pleistocene so the focus on these groups is appropriate given past research goals, but global models should consider the large number of species excluded given emphasis on these four groups, especially given the purported utility of ecometrics as a taxon-free method (J.T. Eronen et al. 2010b; Polly and Head 2015; Vermillion et al. 2018).

Recommendations for future ecometric studies

Despite the mismatch between the estimated and true values for these climatic variables, we believe that dental ecometric approaches represent an extremely valuable tool for environmental reconstruction, given the high preservation potential and information content of large mammal teeth (J.T. Eronen et al. 2010b; Vermillion et al. 2018). We have found that the incompleteness of our understanding of modern mammal communities has limited impact on the accuracy of dental ecometric models. However, the previous focus on reconstructing early hominin environments has led to a lack of utility to previous models elsewhere in the world. We therefore recommend three main changes to previous approaches.

First, given that most regions worldwide contain large mammalian herbivores beyond the four taxa analysed previously, we suggest that future models based on complete large herbivore communities should expand their taxonomic range, to include any species reaching a given

size threshold (e.g., 1 kg) and for which vegetation forms most of their diet. Increasing the taxonomic range of the definition of a large herbivore would allow inclusion of, for example, rodents, carnivorans, hyraxes, xenarthrans (folivorans and extinct cingulates) and marsupials, among others, therefore allowing for application of models to regions, such as Australia and South America that remain understudied in this area. In North America, hypsodonty in Glires has been shown to be an effective predictor of climatic variables including MAT (Schap et al. 2021), supporting the potential utility of dental ecometrics outside of the four taxa used previously. To expand taxon sampling, however, further testing of the functional relationship between the dental traits of these taxa and their environments is required.

Second, we suggest that in most cases, regional-scale models between environmental variables and dental traits should be preferred to global-scale models. This would allow for optimisation of parameters on smaller scale, and reduce the challenge of fitting models to both tropical and temperate latitudes (Liu et al. 2012; Žliobaite et al. 2018; Oksanen et al. 2019). It might also allow inclusion of taxa outside those used previously. For example, in South America, continent-level models may consider giant rodents (e.g., *Josephoartigasia monesi* (Rinderknecht & Blanco 2008)), xenarthrans and South American Native Ungulates that are not relevant in other continents. Global approaches may be most appropriate for some research questions, but we consider regional or continental-scale models to be preferable in most cases.

Finally, we believe that future models should be built based on historical relationships between dental traits and the environment recorded in the fossil record. Previous studies have found that the relationships between traits and the environment differ in contemporary systems compared to historical systems as a result of human effects on ecosystems (Polly 2010; Short et al. 2021a,b; Short et al. 2021b). To eliminate these anthropogenic impacts, we propose reconstructing relationships between traits and the environment based on proxies from the fossil record. This has yet to be trialled, but we believe that it may provide a better method of measuring ecometric load (the mismatch between the real and optimal community trait values (Polly and Head 2015)) due to anthropogenic activity, which is one goal of conservation palaeobiology (Barnosky et al. 2017). Our results suggest that completeness for each community is not essential for obtaining stable estimates (with only ~40% completeness required, assuming the most common mammals are included) (Figure 4). Whilst there are significant gaps in the fossil record, for most large mammals and for more recent time periods, this should be achievable (Zliobaite and Fortelius 2021). We acknowledge that further back in time, relationships between particular traits and the environment are likely to become more different, for example, hypsodonty only evolves in placental taxa in the Miocene, so models based on hypsodonty will not be meaningful beyond this (Damuth and Janis 2011). However, we consider that most trait–environment relationships are likely to hold through deep time and the accuracy of our models will be increased by excluding any anthropogenic effects.

Conclusions

We found that understanding of the taxonomic diversity of large herbivorous mammals remains incomplete and estimate that there are likely to be at least 83 undescribed species. The proportion of undescribed species differs across taxa and space, with the largest number of undescribed species likely to be primates, artiodactyls and marsupials and within primarily tropical biogeographic realms (Neotropical, Afrotropic and Indomalaya, respectively).

Despite the gaps in our understanding of the modern fauna, reconstructions of climate based on dental ecometrics tend to be relatively stable to completeness, above a threshold of ~40% and provided the most common taxa are sampled. However, the large mismatch between true values of environmental variables and our predictions based on dental traits suggests that these models are affected by other confounding factors. To improve the accuracy of predictions based on dental traits, we suggest three possible improvements to approaches that have been used previously (depending on the specific research question):

(1) Community-level models should include all large herbivores within a community, rather than using only limited taxa.
(2) Where appropriate, models should be built on a regional or continental scale to allow optimisation of parameters.
(3) To avoid the effect of anthropogenic impacts on trait–environment relationships, we suggest building models based on the relationships recorded in the fossil record.

Acknowledgements

We are grateful to two anonymous reviewers for their helpful comments in improving a previous version of this manuscript. We thank Michael Mechenich for providing the hexagonal grid cells used for ecometric analyses and the authors of Liu et al. 2012; Oksanen et al. 2019 for making their data publicly available. We are extremely grateful to Indrė Žliobaitė, Tegan Foister and Rob Beckett for their comments on drafts of this manuscript and also thank Vanessa Williams and Hal Wilson for their illustrations and assistance in producing Figures 1 and 2.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Academy of Finland grant NEPA - Non-analogue ecosystems in the past [340775/346292].

ORCID

Oscar E. Wilson http://orcid.org/0000-0002-4046-3128
Mikael Fortelius http://orcid.org/0000-0002-4851-783X
Juha Saarinen http://orcid.org/0000-0003-1351-4652

References

Andrews P, Hixson S. 2014. Taxon-free methods of palaeoecology. Ann Zool Fennici. 51(1–2):269–284.

Barnosky AD, Hadly EA, Gonzalez P, Head J, Polly PD, Lawing AM, Eronen JT, Ackerly DD, Alex K, Biber E, et al. 2017. Merging paleobiology with conservation biology to guide the future of terrestrial ecosystems. Science. 355 (6325):eaah4787. . doi:10.1126/science.aah4787.

Barnosky AD, Koch PL, Feranec RS, Wing SL, Shabel AB. 2004 October. Assessing the causes of late Pleistocene Extinctions on the Continents. Science. 306(5693):70–76. . doi:10.1126/science.1101476.

Barnosky AD, Matzke N, Tomiya S, Wogan GOU, Swartz B, Quental TB, Marshall C, McGuire JL, Lindsey EL, Maguire KC, et al. 2011. Has the Earth's sixth mass extinction already arrived? Nature. 471(7336):51–57. doi:10.1038/nature09678.

Bebber DP, Marriott FHC, Gaston KJ, Harris SA, Scotland RW. 2007. Predicting unknown species numbers using discovery curves. Proc R Soc B Biol Sci. 274 (1618):1651–1658. doi:10.1098/rspb.2007.0464.

Beerling DJ, Lomax BH, Royer DL, Upchurch GR, Kump LR. 2002. An atmospheric pCO2 reconstruction across the cretaceous-tertiary boundary from leaf megafossils. Proc Natl Acad Sci U S A. 99(12):7836–7840. doi:10.1073/pnas.122573099.

Bivand R, Pebesma E, Gomez-Rubio V. 2013. Applied spatial data analysis with R. [place unknown].

Boehm MMA, Cronk QCB. 2021. Dark extinction: the problem of unknown historical extinctions. Biol Lett. 17(3). doi:10.1098/rsbl.2021.0007.
Brown JH, Svenning J-C. 2014. Why are there so many species in the tropics? J Biogeogr. 41(1):8–22. doi:10.1111/jbi.12228.
Burney DA, Ramilisonina. 1998. The Kilopilopitsofy, Kidoky, and Bokyboky: accounts of Strange Animals from Belo-sur-mer, Madagascar, and the Megafaunal "Extinction Window". Am Anthropol. 100(4):957–966. doi:10.1525/aa.1998.100.4.957.
Clarke DA, York PH, Rasheed MA, Northfield TD. 2017. Does biodiversity–ecosystem function literature neglect tropical ecosystems? Trends Ecol Evol. 32(5):320–323. doi:10.1016/j.tree.2017.02.012.
Collen B, Ram M, Zamin T, McRae L. 2008. The tropical biodiversity data gap: addressing disparity in global monitoring. Trop Conserv Sci. 1(2):75–88. doi:10.1177/194008290800100202.
Cozzuol MA, Clozato CL, Holanda EC, Rodrigues FHG, Nienow S, De Thoisy B, Redondo RAF, Santos FR. 2013. A new species of tapir from the Amazon. J Mammal. 94(6):1331–1345. doi:10.1644/12-MAMM-A-169.1.
Cozzuol MA, De Thoisy B, Fernandes-Ferreira H, Rodrigues FHG, Santos FR. 2014. How much evidence is enough evidence for a new species? J Mammal. 95(4):899–905. doi:10.1644/14-MAMM-A-182.
Damuth J, Janis CM. 2011. On the relationship between hypsodonty and feeding ecology in ungulate mammals, and its utility in palaeoecology. Biol Rev. 86 (3):733–758. doi:10.1111/j.1469-185X.2011.00176.x.
Dung VV, Giao PM, Chinh NN, Tuoc D, Arctander P, MacKinnon J. 1993. A new species of living bovid from Vietnam. Nature. 363(6428):443–445. doi:10.1038/363443a0.
Eronen JT, Liu L, Lintulaakso K, Damuth J, Janis CM, Fortelius M, Puolamäki K. 2010a. Precipitation and large herbivorous mammals I: estimates from present-day communities. Evol Ecol Res. 12:217–233.
Eronen JT, Polly PD, Fred M, Damuth J, Frank DC, Scheidegger C, Stenseth NC, Fortelius M, FORTELIUS M. 2010b. Ecometrics : the traits that bind the past and present together. Integr Zool. 5(2):88–101. doi:10.1111/j.1749-4877.2010.00192.x.
Etten JV, Sumner M, Cheng J, Baston D, Bevan A, Bivand R, Busetto L, Canty M, Fasoli B, Forrest D, et al. 2021. Package 'raster'.
Fennessy J, Bidon T, Reuss F, Kumar V, Elkan P, Nilsson MA, Vamberger M, Fritz U, Janke A. 2016. Multi-locus analyses reveal four Giraffe species instead of one. Curr Biol. 26(18):2543–2549. Internet. doi:10.1016/j.cub.2016.07.036.
Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int J Climatol. 37(12):4302–4315. doi:10.1002/joc.5086.
Flannery TF,B, Szalay AL. 1995. A new tree-kangaroo (Dendrolagus : Marsupialia) from Irian Jaya, Indonesia, with notes on ethnography and the evolution of tree-kangaroos. Mammalia. 59(1):65–84.
Foister T, Felice R. 2021. Ecometric modelling of limb proportions and vegetation index among non-human primates in South America. HYSTRIX - the Italian Journal of Mammalogy 1–5 0394–1914 .
Fortelius M, Eronen JT, Jernvall J, Liu L, Pushkina D, Rinne J, Tesakov A, Vislobokova I, Zhang Z, Zhou L. 2002. Fossil mammals resolve regional patterns of Eurasian climate change over 20 million years. Evol Ecol Res. 4 (7):1005–1016.
Fortelius M, Žliobaitė I, Kaya F, Bibi F, Bobe R, Leakey L, Leakey M, Patterson D, Rannikko J, Werdelin L. 2016. An ecometric analysis of the fossil mammal record of the Turkana basin. Philos Trans R Soc B Biol Sci. 371 (1698):20150232. doi:10.1098/rstb.2015.0232.
Freudenthal M, Martín-Suárez E. 2013. Estimating body mass of fossil rodents. Scr Geol. 8(145):1–130.
Fulwood EL. 2020. Ecometric modelling of tooth shape and precipitation gradients among lemurs on Madagascar. Biol J Linn Soc. 129(1):26–40.
Giam X, Scheffers BR, Sodhi NS, Wilcove DS, Ceballos G, Ehrlich PR. 2011. Reservoirs of richness: least disturbed tropical forests are centres of undescribed species diversity. Proc R Soc B Biol Sci. 279(1726):67–76. doi:10.1098/rspb.2011.0433.
Groves CP, Fernando P, Robovský J, DeSalle R. 2010. The sixth rhino: a taxonomic re-assessment of the critically endangered northern white rhinoceros. PLoS One. 5(4):e9703. doi:10.1371/journal.pone.0009703.
Groves CP, Grubb P. 2000. Do Loxodonta cyclotis and L. africana interbreed? Elephant. 2(4):4–7. doi:10.22237/elephant/1521732173.
Groves C, Grubb P. 2011. Ungulate Taxonomy. Baltimore (Maryland): Johns Hopkins University Press.
Grubb P, Groves CP, Dudley JP, Shoshani J. 2000. Living African elephants belong to two species: loxodonta africana (Blumenbach, 1797) and Loxodonta cyclotis (Matschie, 1900). Elephant. 2(4):1–4. doi:10.22237/elephant/1521732169.
Gutiérrez EE, Garbino GST. 2018. Species delimitation based on diagnosis and monophyly, and its importance for advancing mammalian taxonomy. Zool Res. 39(5):301–308.
Head JJ, Bloch JI, Hastings AK, Bourque JR, Cadena EA, Herrera FA, Polly PD, Jaramillo CA. 2009. Giant boid snake from the Palaeocene neotropics reveals hotter past equatorial temperatures. Nature. 457(7230):715–717. Internet. doi:10.1038/nature07671.
Heuvelmans B. 1982. What is Cryptozoology? Cryptozoology. 1:1–12.
Holmes G, Smith TA, Ward C. 2018. Fantastic beasts and why to conserve them: animals, magic and biodiversity conservation. Oryx. 52(2):231–239. doi:10.1017/S003060531700059X.
Jenkins PD, Kilpatrick WC, Robinson MF, Timmins RJ. 2005. Morphological and molecular investigations of a new family, genus and species of rodent (Mammalia: Rodentia: Hystricognatha) from Lao PDR. Syst Biodivers. 2 (4):419–454. doi:10.1017/S1477200004001549.
Jernvall J, Fortelius M. 2002. Common mammals drive the evolutionary increase of hypsodonty in the neogene. Nature. 417(6888):538–540. doi:10.1038/417538a.
Jernvall J, Hunter JP, Fortelius M. 1996 November. Molar tooth diversity, disparity, and ecology in Cenozoic ungulate radiations Jukka. Science. 274 (5292):1489–1492. (80-). doi:10.1126/science.274.5292.1489.
Johnson CN. 2002. Determinants of loss of mammal species during the Late Quaternary "megafauna" extinctions: life history and ecology, but not body size. Proc R Soc B Biol Sci. 269(1506):2221–2227. doi:10.1098/rspb.2002.2130.
Jones KE, Bielby J, Cardillo M, Fritz SA, O'Dell J, Orme CDL, Safi K, Sechrest W, Boakes EH, Carbone C, et al. 2009. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. Ecology. 90(9):2648. doi:10.1890/08-1494.1.
Jones T, Ehardt CL, Butynski TM, Davenport TRB, Mpunga NE, Machaga SJ, De Luca DW. 2005. The highland mangabey Lophocebus kipunji: a new species of African monkey. Science. 308(5725):1161–1164. (80-). doi:10.1126/science.1109191.
Kitchener AC, Breitenmoser-Wursten C, Eizirik E, Gentry A, Werdelin L, Wilting A, Yamaguchi N. 2007. A revised taxonomy of the Felidae: the final report of the Cat classification task force of the IUCN/ SSC Cat specialist group. [place unknown].
Kitchener AC, Hoffmann M, Yamaguchi N, Breitenmoser-Würsten C, Wilting A. 2022. A system for designating taxonomic certainty in mammals and other taxa. Mamm Biol. 102(1):251–261. Internet. doi:10.1007/s42991-021-00205-3.
Lankester ER. 1901a June. Exhibition of, and remarks upon, two skulls and the skin of the newly-discovered African mammal Okapia johnstoni. Proc Zool Soc London 11. 279–281.
Lankester ER. 1901b November. On Okapia, a new genus of Giraffidae from Central Africa. Trans Zool Soc London 16. 471–474.
Leydesdorff L, Persson O. 2010. Mapping the geography of science: distribution patterns and networks of relations among cities and institutes. J Am Soc Inf Sci Technol. 61(8):1622–1634. Internet. http://onlinelibrary.wiley.com/doi/10.1002/asi.22883/abstract.
Liu L, Puolamäki K, Eronen JT, Ataabadi MM, Hernesniemi E, Fortelius M. 2012. Dental functional traits of mammals resolve productivity in terrestrial ecosystems past and present. Proc R Soc B Biol Sci. 279(1739):2793–2799. doi:10.1098/rspb.2012.0211.
Liu J, Slik F, Zheng S, Lindenmayer DB. 2022 February. Undescribed species have higher extinction risk than known species. Conserv Lett e12876. 1–8.
Matschie P. 1900. Über geographische Albarten des Afrikanischen elephantens. Sitzungsber der Ges Naturforsch Freunde zu Berlin. 8:189–197.
Meloro C, Kovarovic K, Veech J. 2013. Spatial and ecometric analyses of the Plio-Pleistocene large mammal communities of the Italian peninsula. J Biogeogr. 40(8):1451–1462. doi:10.1111/jbi.12113.
Menegotto A, Rangel TF. 2018. Mapping knowledge gaps in marine diversity reveals a latitudinal gradient of missing species richness. Nat Commun. 9(1). Internet. doi:10.1038/s41467-018-07217-7.
Meyer M, Palkopoulou E, Baleka S, Stiller M, Penkman KEH, Alt KW, Ishida Y, Mania D, Mallick S, Meijer T, et al. 2017. Palaeogenomes of Eurasian straight-tusked elephants challenge the current view of elephant evolution. Elife. 6:1–14. doi:10.7554/eLife.25413.
Mill Valley Independent. 1909. Johnston and the okapi: how Sir Harry, recent guest of Roosevelt, made his remarkable find. Mill Val Indep. 11 (29).
Mittelbach GG, Schemske DW, Cornell HV, Allen AP, Brown JM, Bush MB, Harrison SP, Hurlbert AH, Knowlton N, Lessios HA, et al. 2007. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. Ecol Lett. 10(4):315–331. doi:10.1111/j.1461-0248.2007.01020.x.
Nater A, Mattle-Greminger MP, Nurcahyo A, Nowak MG, de Manuel M, Desai T, Groves C, Pybus M, Sonay TB, Roos C, et al. 2017. Morphometric, behavioral, and genomic evidence for a new orangutan species. Curr Biol. 27 (22):3487–3498.e10. doi:10.1016/j.cub.2017.09.047.

Negri A, Amorosi A, Antonioli F, Bertini A, Florindo F, Lurcock PC, Marabini S, Mastronuzzi G, Regattieri E, Rossi V, et al. 2015. A potential global boundary stratotype section and point (GSSP) for the tarentian stage, upper Pleistocene, from the Taranto area (Italy): results and future perspectives. Quat Int. 383:145–157. Internet. doi:10.1016/j.quaint.2014.08.057.

Oates JF, Woodman N, Gaubert P, Sargis EJ, Wiafe ED, Lecompte E, Dowsett-Lemaire F, Dowsett RJ, Gonedelé Bi S, Ikemeh RA, et al. 2021 June. A new species of tree hyrax (Procaviidae: Dendrohyrax) from West Africa and the significance of the Niger–Volta interfluvium in mammalian biogeography. Zool J Linn Soc 194 (2): 1–26 doi:10.1093/zoolinnean/zlab029.

Oksanen O, Žliobaitė I, Saarinen J, Lawing AM, Fortelius M. 2019. A Humboldtian approach to life and climate of the geological past: estimating palaeotemperature from dental traits of mammalian communities. J Biogeogr. 46(8):1760–1776. doi:10.1111/jbi.13586.

Oren DC. 2001. Does the Endangered Xenarthran Fauna of Amazonia include remnant ground sloths? Edentata . 4:2–5.

Padovani V, Rossi L. 2021. The science of monster hunting. Atti dela Soc dei Nat e Mat di Modena. 152:141–153. https://www.researchgate.net/publication/355588200.

The Paleobiology Database. 2021. The Paleobiology Database [Internet]. [accessed 2020 Aug 25]. https://paleobiodb.org.

Patterson BD. 2000. Patterns and trends in the discovery of new Neotropical mammals. Divers Distrib. 6(3):145–151. doi:10.1046/j.1472-4642.2000.00080.x.

Pebesma E, Bivand R. 2005. Classes and methods for spatial data in R. R News. 5 (2):9–13.

Pfenninger M, Schwenk K. 2007. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. BMC Evol Biol. 7 (1):1–6. doi:10.1186/1471-2148-7-121.

Pine RH. 1994 April. New mammals not so seldom. Nature. 368(6472):1994. doi:10.1038/368593a0.

Pineda-Munoz S, Wang Y, Kathleen Lyons S, Tóth AB, McGuire JL. 2021. Mammal species occupy different climates following the expansion of human impacts. Proc Natl Acad Sci U S A. 118(2). doi:10.1073/pnas.1922859118.

Polly PD. 2010 Tiptoeing through the trophics: geographic variation in carnivoran locomotor ecomorphology in relation to environment Goswami, A, Frischia, A eds.Carnivoran Evolution: New Views on Phylogeny, Form, and Function, Cambridge University Press: Cambridge, UK. 374–410.

Polly PD, Head JJ. 2015. Measuring earth-life transitions: ecometric analysis of functional traits from late Cenozoic vertebrates. Paleontol Soc Pap. 21:21. doi:10.1017/S1089332600002953.

Pujos F, De Iuliis G, Cartelle C. 2016. A paleogeographic overview of tropical fossil sloths: towards an understanding of the origin of extant suspensory sloths? J Mamm Evol. 24(1):19–38. Internet. doi:10.1007/s10914-016-9330-4.

Raja NB, Dunne EM, Matiwane A, Khan TM, Nätscher PS, Ghilardi AM, Chattopadhyay D. 2021 Colonial history and global economics distort our understanding of deep-time biodiversity. Nature Ecology and Evolution. 6 145–154 doi:10.1038/s41559-021-01608-8.

Rapoport E. 1982. Areography: geographical strategies of species. [place unknown]. Oxford: Pergamon.

Reeder DM, Helgen KM, Wilson DE. 2007. Global Trends and Biases in New Mammal Species Discoveries. Occas Pap Museum Texas Tech Univ. 269:1–35.

Rinderknecht A, Blanco RE. 2008. The largest fossil rodent. Proc R Soc B Biol Sci. 275(1637):923–928. doi:10.1098/rspb.2007.1645.

Ritz C, Baty F, Streibig JC, Gerhard D, Xia Y. 2015. Dose-response analysis using R. PLoS One. 10(12):1–13. doi:10.1371/journal.pone.0146021.

Roca AL, Georgiadis N, Pecon-Slattery J, O'Brien SJ. 2001. Genetic evidence for two species of elephant in Africa. Science. 293(5534):1473–1477. (80-). doi:10.1126/science.1059936.

Roca AL, Ishida Y, Brandt AL, Benjamin NR, Zhao K, Georgiadis NJ. 2015 November. Elephant natural history: a genomic perspective. Annu Rev Anim Biosci. 3(1):139–167. doi:10.1146/annurev-animal-022114-110838.

Rolland J, Condamine FL, Jiguet F, Morlon H, Moritz C. 2014. Faster speciation and reduced extinction in the tropics contribute to the Mammalian Latitudinal diversity gradient. PLoS Biol. 12(1):e1001775. doi:10.1371/journal.pbio.1001775.

Rolland J, Silvestro D, Schluter D, Guisan A, Broennimann O, Salamin N. 2018. The impact of endothermy on the climatic niche evolution and the distribution of vertebrate diversity. Nat Ecol Evol. 2(3):459–464. Internet. doi:10.1038/s41559-017-0451-9.

Rossi L. 2016 A Review of Cryptozoology: Towards a Scientific Approach to the Study of Hidden Animals Angelici, F.M ed. Switzerland: Springer International Publishing. 573–588.

Rossi L, Gippoliti S, Angelici FM. 2018. The role of indirect evidence and traditional ecological knowledge in the discovery and description of new ape and monkey species since 1980. Primates. 59(4):327–337. doi:10.1007/s10329-018-0667-6.

Rosti H, Pihlström H, Bearder S, Pellikka P, Rikkinen J. 2020. Vocalization analyses of Nocturnal Arboreal Mammals of the Taita Hills, Kenya. Diversity. 12(473):1–23. doi:10.3390/d12120473.

RStudioTeam. 2021. RStudio: integrated development for R [Internet]. http://www.rstudio.com/.

Saarinen J, Eronen JT, Fortelius M, Seppä H, Lister AM. 2016. Patterns of diet and body mass of large ungulates from the pleistocene of Western Europe, and their relation to vegetation. Palaeontol Electron. 19 (3): 1–58.

Saarinen J, Oksanen O, Žliobaitė I, Fortelius M, DeMiguel D, Azanza B, Bocherens H, Luzón C, Solano-García J, Yravedra J, et al. 2021. Pliocene to Middle Pleistocene climate history in the Guadix-Baza Basin, and the environmental conditions of early Homo dispersal in Europe. Quat Sci Rev. 268:107132. doi:10.1016/j.quascirev.2021.107132.

Schaller GB, Vrba ES. 1996. Description of the giant muntjac (Megamuntiacus vuquangensis) in Laos. J Mammal. 77(3):675–683. doi:10.2307/1382672.

Schap JA, Samuels JX, Joyner TA. 2021 [Internet]. 562(November 2020). Ecometric estimation of present and past climate of North America using crown heights of rodents and lagomorphs. Palaeogeogr Palaeoclimatol Palaeoecol. 110144. doi:10.1016/j.palaeo.2020.110144.

Sclater PL. 1901a On an apparently new Species of Zebra from the Semliki Forest In: Proc Zool Soc 11: 50–52.

Sclater PL. 1901b May. Exhibition of water-colour painting of okapi and extracts of a letter written by Sir Harry Johnston. Proc Zool Soc London 11: 2–6.

Short RA, Lawing AM, Thomassen H. 2021a. Geography of artiodactyl locomotor morphology as an environmental predictor. Divers Distrib. 27 (9):1818–1831. doi:10.1111/ddi.13371.

Short RA, Pinson K, Lawing AM. 2021b. Comparison of environmental inference approaches for ecometric analyses: using hypsodonty to estimate precipitation. Ecol Evol. 11(1):587–598. doi:10.1002/ece3.7081.

Shuker KPN. 2012. The Encyclopedia of New and Rediscovered Animals. Landisville (Pennsylvania): Coachwhip Publications.

Shuker KPN. 2016. Still in search of prehistoric survivors. Greenville (Ohio): Coachwhip Publications.

Stevens GC. 1989. The latitudinal gradient in geographical range: how so many species coexist in the tropics. Am Nat. 133(2):240–256. doi:10.1086/284913.

Stroud JT, Feeley KJ. 2017. Neglect of the tropics is widespread in ecology and evolution: a comment on Clarke et al. Trends Ecol Evol. 32(9):626–628. Internet. doi:10.1016/j.tree.2017.06.006.

van Roosmalen MGM. 2013a. New species of living tapir, the dwarf tapir (Mammalia: Tapiridae) from the Brazilian Amazon. In: Barefoot through Amazon - On the Path of Evolution. North Charleston SC (USA): CreateSpace; p. 400–404.

van Roosmalen MGM. 2013b. Barefoot in the Amazon - on the path of evolution. North Charleston SC (U.S.A): CreateSpace.

van Roosmalen MGM. 2014. Case 3650 Tapirus pygmaeus Van Roosmalen & Van Hooft in Van Roosmalen, 2013 (Mammalia, Perissodactyla, tapiridae): proposed confirmation of availability of the specific name and of the book in which this nominal species was proposed. Bull Zool Nomencl. 71(2):84–87. doi:10.21805/bzn.v71i2.a8.

Velden FFV. 2016. Realidade, ciência e fantasia nas controvérsias sobre o Mapinguari no sudoeste amazônico. Bol Do Mus Para Emilio GoeldiCiencias Humanas. 11(1):209–224. doi:10.1590/1981.81222016000100011.

Vermillion WA, Polly PD, Head JJ, Eronen JT, Lawing AM. 2018 Ecometrics: a trait-based approach to paleoclimate and paleoenvironmental reconstruction Croft, DA, Su, DF, Simpson, SW eds. Methods in Paleoecology: Reconstructing Cenozoic Terrestrial Environments and Ecological Communities Vertebrate Paleobiology and Paleoanthropology. Switzerland: Springer Nature. 373–394.

Vilhena DA, Smith AB, Butler RJ. 2013. Spatial bias in the marine fossil record. PLoS One. 8(10):1–7. doi:10.1371/journal.pone.0074470.

Violle C, Navas M-L, Vile D, Kazakou E, Fortunel C, Hummel I, Garnier E. 2007. Let the concept of trait be functional! Oikos. 116(5):882–892. doi:10.1111/j.0030-1299.2007.15559.x.

Voss RS, Helgen KM, Jansa SA. 2014. Extraordinary claims require extraordinary evidence: a comment on Cozzuol et al. (2013). J Mammal. 95(4):893–898. doi:10.1644/14-MAMM-A-054.

Watson DM, Davis RA. 2017. Hopeful Monsters—In defense of quests to rediscover long-lost species. Conserv Lett. 10(4):381–382. doi:10.1111/conl.12386.
Wetzel RM, Dubos RE, Martin RL, Myers P. 1975. Catagonus, an "Extinct" Peccary, Alive in Paraguay. Science. 189(4200):379–381. (80-). doi:10.1126/science.189.4200.379.
Wickham H. 2016. ggplot2: elegant graphics for data analysis. place unknown: Springer-Verlag New York.
Woodley MA, Naish D, Shanahan HP. 2009. How many extant pinniped species remain to be described? Hist Biol. 20(4):225–235. doi:10.1080/08912960902830210.
Wozencraft W. 2005. Order Carnivora. In: Wilson DE, Reeder D, editors. Mammal species world A Taxon Geogr Ref. 3rd ed. Baltimore (Maryland): Johns Hopkins University Press; p. 532–628.
Xu X, Arnason U. 1996. The mitochondrial DNA Molecule of Sumatran orangutan and a molecular proposal for two (Bornean and Sumatran) species of orangutan. J Mol Evol. 43(5):431–437. doi:10.1007/BF02337514.
Zamani A, Orfinger AB, Gante HF, Scherz MD, Pos DD. 2022. DNA barcodes on their own are not enough to describe a species. (October 2021). 1–5.
Zliobaite I, Fortelius M. 2021. On calibrating the completometer for the mammalian fossil record. Paleobiology. 1–11. doi:10.1017/pab.2021.22.
Žliobaite I, Rinne J, Tóth AB, Mechenich M, Liu L, Behrensmeyer AK, Fortelius M. 2016. Herbivore teeth predict climatic limits in Kenyan ecosystems. Proc Natl Acad Sci U S A. 113(45):12751–12756. doi:10.1073/pnas.1609409113.
Žliobaite I, Tang H, Saarinen J, Fortelius M, Rinne J, Rannikko J. 2018. Dental ecometrics of tropical Africa: linking vegetation types and communities of large plant-eating mammals. Evol Ecol Res. 19(3):127–147.